Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2023) 31: 1329 – 1342
© TÜBİTAK
doi:10.55730/1300-0632.4050

Feature selection optimization with filtering and wrapper methods: two disease classification cases

Serhat ATİK[*], **Tuğba DALYAN**
Department of Computer Engineering, Faculty of Engineering and Natural Sciences
İstanbul Bilgi University, İstanbul, Turkiye

Received: 05.10.2022 • **Accepted/Published Online:** 12.10.2023 • **Final Version:** 30.11.2023

Abstract: Discarding the less informative and redundant features helps to reduce the time required to train a learning algorithm and the amount of storage required, improving the learning accuracy as well as the quality of results. In this study, we present different feature selection approaches to address the problem of disease classification based on the Parkinson and Cardiac Arrhythmia datasets. For this purpose, first we utilize three filtering algorithms including the Pearson correlation coefficient, Spearman correlation coefficient, and relief. Second, metaheuristic algorithms are compared to find the most informative subset of the features to obtain better classification accuracy. As a final method, a hybrid model involving filtering algorithms is applied to the datasets to eliminate half of the features, and then a metaheuristic algorithm based on a proposed genetic algorithm is applied to the rest of the datasets. With all three methods, we use three classification algorithms: support vector machine, K-nearest neighbor, and random forest. The results show that the best scores are obtained from the metaheuristic algorithm based on the proposed genetic algorithm for both datasets. This comparative study contributes to the literature by increasing the accuracy of classification for both datasets and presenting a hybrid model with filtering and a metaheuristic algorithm.

Key words: Feature selection, optimization algorithms, metaheuristic algorithms, genetic algorithms, filtering methods

1. Introduction

Technological developments and the growing role of artificial intelligence (AI) provide tremendous opportunities and applications across a variety of sectors, such as health care, finance, agriculture, and education. In particular, the use of AI in health care offers a number of advantages over conventional methods and helps support clinical decision-making techniques with the development of better treatments for unmet medical needs, the monitoring of patients, and the revolutionizing of drugs.

Parkinson disease is a common and the fastest growing neurodegenerative movement disorder and more than 10 million people live with this disease today [1]. It is not a totally preventable disease in the early stages, but diagnosing it in the early stages may improve the life quality of the patients. Statistical results show that Parkinson disease increases with age but 4% of patients are less than 50 years old.[1] Such results and predictions play an important role in analyzing how these diseases should be approached and treated. Cardiac arrhythmia is another important disease caused by the heart's disrupted normal rhythm. Similarly to Parkinson disease, detecting cardiac arrhythmia and offering treatment earlier may improve the life quality of the patients.

[*] Correspondence: serhat.atk@outlook.com

[1] Parkinson's Foundation

In this study, we selected two commonly used datasets: the Parkinson (754 features) [2] and Cardiac Arrhythmia (279 features) [3] datasets. The main problem with both datasets is having too many features. To deal with these types of high-dimensional datasets, feature selection is one of the most popular solutions to reduce irrelevant and redundant characteristics from a dataset and use the best subset of features for learning techniques. Discarding the less informative features leads to lower computational cost and better learning performance as well as helping to reduce the time required to train a learning algorithm and the amount of storage required.

Feature selection methods can be roughly classified into five groups: wrapper, filtering, embedded, ensemble, and hybrid models [4–10]. Filtering methods use an indirect criterion to measure the success of the predictions and rely on statistical approaches. Wrapper methods simply create agents in the search space and evaluate those agents with a learning model. There are many differences between filter and wrapper methods. While filter methods are computationally faster, avoid overfitting, and may sometimes fail to select the best features, wrapper methods use predictive models, are slower than filter methods and prone to overfitting, and show better performance in general. On the other hand, embedded methods are frequently assigned to learners and perform feature selection during the learning process. These models build on earlier models by employing different evaluation criteria at different stages of the search. They are less computationally expensive than wrapper methods and less prone to overfitting. Ensemble methods are especially useful to get rid of the local optimum by combining multiple feature subsets to select an optimal subset of features using a combination of feature rankings that improves classification accuracy. Finally, hybrid models are combinations of these previous models.

In this study, the first aim is to use different types of feature selection algorithms to eliminate noise or redundant features and compare the results of the filtering and wrapper methods based on metaheuristic algorithms. The second main aim is to observe the contribution of a hybrid method using filtering and a metaheuristic algorithm based on a proposed genetic algorithm (GA) and draw comparisons to previous studies [3, 11]. The final aim is to use three classification algorithms, namely support vector machine (SVM), k-nearest neighbor (kNN), and random forest (RF), to address the problems of classifying the diseases. The algorithms used in this study are listed in Table 1. The main contributions of this paper are as follows:

- Proposing a hybrid model using filtering algorithms and a metaheuristic algorithm based on a proposed GA.
- Comparing metaheuristic algorithms among themselves.
- Confirming the better performance of the promising results of our proposed GA compared to previous studies [3, 11] using the same datasets.

The rest of this paper is organized as follows: Section 2 presents preliminaries on feature selection. The methodology is described in Section 3. Experimental results and a discussion are presented in Section 4. Conclusions and future research directions are given in Section 5.

2. Related work

There are many studies on feature selection methods in the literature. Agrawal et al. [12] presented a comprehensive survey about metaheuristic algorithms for the feature selection problem. Their study comprised a literature review of studies published between 2009 and 2019. They introduced four types of metaheuristic algorithms: evolution-based algorithms, physics-based algorithms, swarm intelligence-based algorithms, and human

Table 1. Filtering and wrapper algorithms used in this study.

Filtering methods	**Wrapper methods**
Pearson Correlation Coefficient (PCC) Spearman Correlation Coefficient (SCC) Relief	Binary Bat Algorithm (BBA) Cuckoo Search Algorithm (CSA) Equilibrium Optimizer (EO) Genetic Algorithm (GA) Gravitational Search Algorithm (GSA) Grey Wolf Optimizer (GWO) Harmony Search (HS) Mayfly Algorithm (MA) Particle Swarm Optimizer (PSO) Red Deer Algorithm (RDA) Sine Cosine Algorithm (SCA) Whale Optimization Algorithm (WOA)

behavior-related algorithms. The survey concluded with a case study based on the University of California-Irvine (UCI) Repository. Another recent survey [13] reviewed and analyzed widely known metaheuristic feature selection algorithms according to some criteria and indicators such as performance, process, practices, challenges, and datasets. Finally, another study [14] presented a comprehensive review of ten different metaheuristic algorithms to address the feature selection problem. Twelve real-valued and mostly high-dimensional UCI benchmark datasets were utilized to compare the performances of these algorithms.

In addition to surveys, various studies have focused on the feature selection problem. Enireddy et al. [11] studied the same Parkinson disease dataset that is used in our study. First, they used a min-max scalar and proposed two different GWO algorithms as metaheuristic algorithms to optimize the feature selection. After the feature selection process, they experimented with machine learning algorithms such as the extreme gradient boosting (XGB) classifier, kNN, and SVM. They reported that the best accuracy was 97.96% with the XGB classifier. Guvenir et al. [3] presented the Cardiac Arrhythmia dataset that is used in our study and proposed a novel machine learning technique named voting feature intervals (VFI5) for diagnosing cardiac arrhythmia using normal 12-lead ECG records. VFI5 is a supervised and inductive learning technique that uses examples to infer categorization expertise. A training set of records that includes clinical measures derived from ECG signals as well as other information such as age, sex, and weight and an expert cardiologist's conclusion is utilized. The knowledge representation is based on a new approach known as feature intervals, in which an idea is represented by the projections of the training examples on each feature individually. With VFI5, classification is determined by a majority vote among the class predictions generated by each feature individually. The authors obtained 62% accuracy with the proposed algorithm and increased that rate to 68% using weights learned by a GA.

Oreski et al. [15] studied a hybrid GA with neural networks for the optimization of credit risk classification. The authors stated that nature-inspired algorithms are simply used with a neural network classification algorithm as a fitness function and they measured the success of the agents. The results showed that feature selection improves the classification accuracy and performance of the hybrid GA better than a simple GA. Alyasiri et al. [16] used wrapper and hybrid feature selection methods to classify English texts. The results showed that text classification accuracy has successful results with metaheuristic algorithms. Abdollahi et al. [17] used a hybrid method to diagnose heart disease, employing feature selection based on an ensemble classifier.

A GA was used to eliminate the redundant features. The authors remarked that the model was improved with the ensemble classifier method and the results were vital for patients.

Sayed et al. [18] proposed a new hybrid algorithm, namely the binary clonal flower pollination algorithm with optimum-path forest classifier as an objective function. The results were compared with well-known algorithms such as the BBA and CSA on three UCI datasets and showed significant performance among other algorithms. Arora et al. [19] proposed a hybrid algorithm based on GWO with the crow search algorithm. The authors compared the hybrid solution with ten other algorithms to prove its success on 23 different benchmark datasets. The findings indicated that the proposed algorithm had remarkable results in terms of optimization ability and computational cost. Yet another study [20] used a metaheuristic algorithm called ant colony optimization based on semisupervised feature selection by measuring relevancy in a supervised method and redundancy in an unsupervised method. The authors used a nonlinear heuristic function rather than a linear one and concluded that the results showed high accuracy with reasonable execution time using 14 benchmark datasets and eight semisupervised feature selection methods.

Hancer et al. [21] proposed a multiobjective (MO) artificial bee colony (ABC) algorithm integrated with a nondominated sorting procedure and genetic operators to optimize the feature selection process for 12 different benchmark datasets. Two different approaches were considered: binary and continuous representation. The authors demonstrated that the performance of binary representation had the best results among other models. Al-Tashi et al.[22] proposed a binary MO GWO based on a sigmoid transfer function to solve feature selection problems. The aim of this paper was to find the best trade-offs between the selected features and error rate. The proposed algorithm was compared with the native MO GWO on 15 different datasets. Another study [23] proposed MO feature selection based on the forest optimization algorithm on nine different UCI datasets. The authors approached two different forest optimization algorithms including continuous and binary representations with a divide-and-conquer approach. The continuous version showed significant performance in addressing the problem of feature selection. A similar approach [24] was proposed with the MO Harris hawks optimization algorithm with different search algorithms including associative learning, chaotic local search, and GWO to eliminate irrelevant features using 16 UCI datasets.

Nouri-Moghaddam et al. [25] proposed a hybrid filter-wrapper solution based on the Fisher score method and MO forest optimization algorithm not only to optimize feature selection but also for optimizing the kernel parameters in SVM classification. The authors compared their results with four different algorithms on six datasets and indicated that the proposed solution outperformed the other four methods. Another study [26] on the feature selection problem for gene selection proposed a framework to extract prominent genes from ten cancer gene expression datasets using two stages: implementation of a weighted gene co-expression network with RF and minimal redundancy maximal relevance for feature selection with the proposal of a binary salp swarm algorithm with machine learning methods for new gene selection. The authors compared the proposed framework with five different optimization algorithms and showed that it had accuracy of over 97.6%.

3. Methodology

In this study, the Parkinson Disease (754 features) and Cardiac Arrhythmia (279 features) datasets are selected from the UCI Data Repository.[2] The main motivation in the selection of both datasets is to draw comparisons with previous studies and to analyze whether our proposed method contributes to the results or not. A general

[2] UCI Data Repository

overview of this study is summarized in Figure 1. The feature selection process starts with a data preprocessing step. We then utilize a filtering method using three algorithms: Pearson correlation coefficient (PCC), Spearman correlation coefficient (SCC), and relief. We also present a wrapper method using 12 metaheuristic algorithms. The final method is a hybrid model in which filtering models are implemented on the datasets to eliminate half of the features to find the best subset of features and a metaheuristic algorithm based on a proposed GA is implemented on the rest of the datasets. Finally, the features obtained from all three proposed methods are used with classification algorithms such as the kNN, SVM, and RF to evaluate the accuracy.

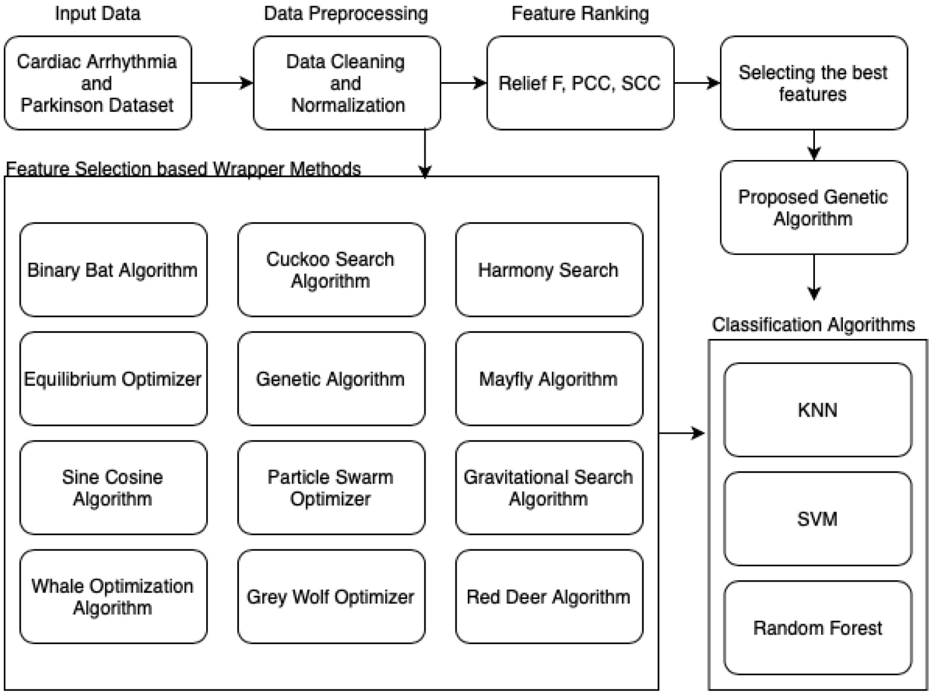


Figure 1. Structure of feature selection.

3.1. Data

The Parkinson dataset was collected from the Department of Neurology of the Cerrahpaşa Faculty of Medicine, İstanbul University. The dataset includes 756 instances and 754 attributes. The instances were collected from 188 patients and 64 healthy individuals. Patients had already been diagnosed with Parkinson disease and their ages were between 33 and 87. On the other hand, healthy individuals' ages were between 41 and 82. The attributes are related to the numeric data of voices and the class of the dataset is binary, showing that an individual has Parkinson disease or not. While collecting the data, the microphone was set to 44.1 kHz and each individual was asked to repeat the sustained phonation of the vowel /a/ three times [2].

The Cardiac Arrhythmia dataset consists of 452 patient records for 279 features. There are 16 different classes, whereby 01 refers to normal ECG, 02 to ischemic changes (coronary artery disease), 03 to old anterior myocardial infarction, 04 to old inferior myocardial infarction, 05 to sinus tachycardia, 06 to sinus bradycardia, 07 to ventricular premature contraction, 08 to supraventricular premature contraction, 09 to left bundle branch block, 10 to right bundle branch block, 11 to first-degree atrioventricular block, 12 to second-degree atrioventricular block, 13 to third-degree atrioventricular block, 14 to left ventricular hypertrophy, 15 to atrial fibrillation or flutter, and 16 to the rest. Features include some personal information such as age, sex, height, and weight. A large portion of the features are the numeric data of the ECG results [3].

3.2. Methods

The collected datasets were preprocessed in several steps using text-processing libraries. The Parkinson dataset was clear; therefore, we only normalized the data using a min-max scalar. On the other hand, the Cardiac Arrhythmia dataset had missing values; therefore, missing values were filled with the "interpolate" function. After that, the min-max scalar was applied for the normalization of the data. The data were also split into training and testing datasets. In the case of holdout cross-validation, the training dataset included 70% of the total dataset and separation was done with a constant random seed value of ten.

As a first method, three different filter-based algorithms including PCC, SCC, and relief were used to analyze the performance of the filtering algorithms. Filter-based feature selection methods do not use any intermediate learning algorithms. Instead, they employ statistical methods to determine the relative relevance of the features in context. The features are ranked by the filter-based methods and then the top features can be used for classification algorithms. The PCC is a linear correlation measurement between two data points. The ratio shows the covariance of two variables and the product of their standard deviations. The ratio is between –1 and 1. If the ratio is close to 1, it means that the relation is strong. The SCC describes the relationship of the features with a monotonic function. The main difference between the PCC and SCC is that while the PCC is limited to linearity, the SCC looks for monotonic relationships. Finally, relief is another statistical-based feature selection algorithm. A feature score is assigned to a feature and is subsequently used to rank and select the highest-scoring features. Feature scores are also used as weights of features for assisting downstream modeling.

Metaheuristic algorithms are problem-independent frameworks that operate based on previously specified strategies to achieve optimization. Important examples of metaheuristics include GAs, GWO, and PSO. Metaheuristic algorithms are mainly inspired by natural environments and always operate with basic heuristic approaches, although they are customized for different strategies. Thus, they are distinguished from optimum solution approaches, because in exact optimization methods, an optimal solution is reached, and this usually happens after a long computational period.

The benefits of metaheuristics can be observed best in complicated models with heavy data loads because metaheuristics are able to obtain substantially better solutions with less error and relatively short amounts of time. In addition, metaheuristics are flexible and can be customized and fitted to real-life scenarios with great ease. This seems to be another advantage over exact optimization methods, because those methods are generally more rigid. In this study, 12 different metaheuristic algorithms including BBA [27], CSA [28], EO [29], GA [30], GSA [31], GWO [32], HS [33], MA [34], PSO [35], RDA [36], SCA [37], and WOA [38] are selected to be applied to both datasets for feature selection.

The GA is the most critical metaheuristic algorithm in this paper because the best accuracy is gained with the GA for both the Parkinson and Cardiac Arrhythmia datasets. The GA is a population-based metaheuristic method commonly used for optimization and searching problems. It is inspired by the natural selection process, where the fittest individuals are selected for reproduction in order to produce the offspring of the next generation. Thus, these offspring inherit the characteristics of the parents. If the parents have better fitness, their offspring will be better than the parents and will have a better chance at surviving. This process has iterations and new populations have totally higher fitness scores at the end. There are five important steps of the GA: the initial population, fitness function, selection, cross-over, and mutation.

In this paper, 12 different metaheuristic [39] and three different filter-based algorithms are compared. Furthermore, we also propose a GA in this paper. In our GA, the agents' genotypes include features such as binary strings and a random number for the mutation operation. The random number shows the number of features that are changed in the mutation operation. The length of the string is equal to the total number of features in the dataset. In the binary string, "zero" shows that a feature is not included and "one" shows that the feature is included. In the fitness step, the algorithm creates a new subset according to the "one" values of the agent's genotype. For score calculation, it implements the classification algorithm and equalizes the accuracy of the test set to the score of the agent. For the next generations, the algorithm selects two agents as parents. The selection operation is called "roulette wheel selection," in which the selection of highly scored agents is more possible. With those parents, the algorithm creates two child agents, implementing cross-over and mutation. There are two different cross-over functions in the algorithm. The first takes the even-numbered genes from the first parent and odd-numbered genes from the second parent. The second cross-over function does the exact opposite of the first cross-over function. After getting two child agents from those two functions, it implements mutation according to a probability rate. The mutation function changes ones to zeros or zeros to ones in the genotype for randomly selected features. The total number of changed features is related to the genotype of the agent, as mentioned before. Finally, the algorithm returns the best agent in the populations. The flow chart of the proposed GA is demonstrated in Figure 2 and the pseudo-code of the algorithm is shown in Figure 3.

The parameters used in the proposed GA are shown in Table 2. Parameters are gathered with a trial-and-error method. In terms of not increasing the computation time too much, population size and number of generations are selected to be low. Moreover, increasing the population size and number of generations does not drastically affect the results. Cross-over is implemented for each new population to diversify the solution space. The mutation rates are set between 20% and 30% and 20% is selected to increase consistency.

Table 2. Proposed GA parameters.

Parameter name	**Value**
Population size	20
Number of generations	100
Probability of cross-over	1
Probability of mutation	0.2
Type of mutation	One-by-one cross
Type of selection	Rank-based roulette wheel
Elite count	2
Number of variables	Parkinson: 754; Cardiac Arrhythmia: 279
Fitness function	Classification algorithms' accuracy
Encoding	Binary

In this study, the grid search algorithm from the scikit-learn library is used for hyperparameter tuning for machine learning classification algorithms. Grid search is a method of searching a manually defined portion of the hyperparameter space of a given algorithm exhaustively. After the grid search implementation, the SVM, kNN, and RF algorithms are used for classification. These algorithms and their parameters are given in Table 3.

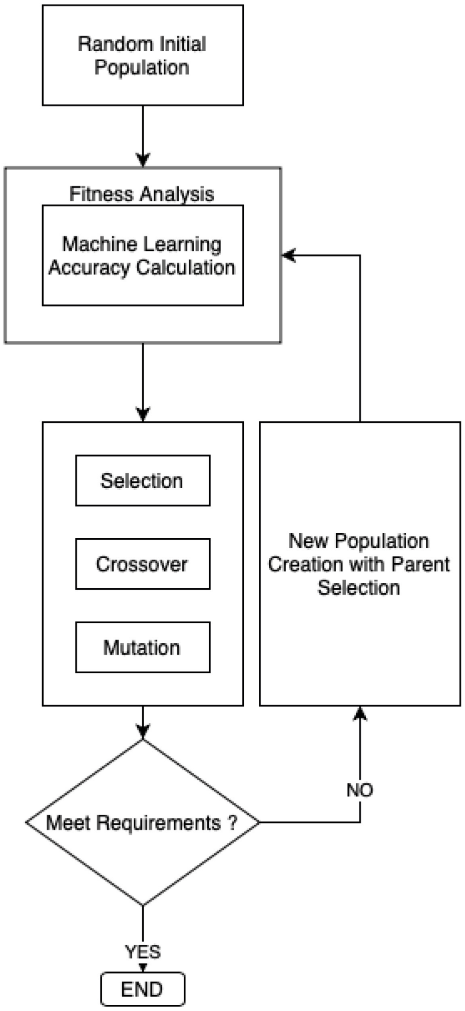
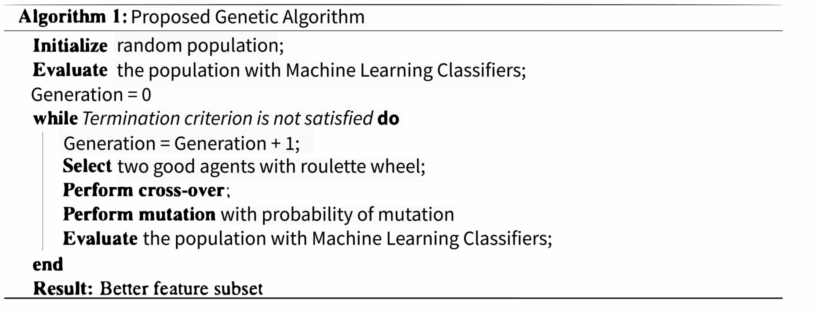

Random Initial Population
Fitness Analysis
Machine Learning Accuracy Calculation
Selection
Crossover
Mutation
New Population Creation with Parent Selection
Meet Requirements ?
NO
YES
END

Figure 2. Flowchart of proposed GA.



```
Algorithm 1: Proposed Genetic Algorithm
  Initialize random population;
  Evaluate the population with Machine Learning Classifiers;
  Generation = 0
  while Termination criterion is not satisfied do
      Generation = Generation + 1;
      Select two good agents with roulette wheel;
      Perform cross-over;
      Perform mutation with probability of mutation
      Evaluate the population with Machine Learning Classifiers;
  end
  Result: Better feature subset
```

Figure 3. Pseudo-code of proposed GA.

4. Experimental results

In the first experiment, filtering algorithms are compared using the Parkinson and Cardiac Arrhythmia datasets. Half of the features are eliminated according to feature importance. Classification algorithms are evaluated for

Table 3. Parameters used after implementing grid search.

kNN		**SVM**		**RF**	
Parkinson	Arrhythmia	Parkinson	Arrhythmia	Parkinson	Arrhythmia
n_neighbors 2	n_neighbors 16	C 10	C 0.1	criterion entropy	criterion gini
metric manhattan	metric manhattan	gamma 0.1	gamma 0.1	max_depth 8	max_depth 8
weights distance	weights distance	kernel rbf	kernel poly	max_features sqrt	max_features auto
				n_estimators 200	n_estimators 200

the rest of the datasets and the accuracies of the algorithms are indicated in Table 4 and Table 5. The results show that while PCC has the best accuracy at 91.19% for Parkinson disease, SCC outperformed the other algorithms at 73.53% for cardiac arrhythmia disease classification. The best results are obtained from the RF algorithm for both datasets. Although the results are obtained within a short time, the accuracies of the algorithms are not sufficient.

Table 4. Results of filtering methods on Parkinson dataset.

	SVM	**kNN**	**RF**
Pearson correlation coefficient	88.55	84.58	**91.19**
Spearman correlation coefficient	84.14	87.67	86.34
Relief	85.90	86.78	90.75

Table 5. Results of filtering methods on Cardiac Arrhythmia dataset.

	SVM	**kNN**	**RF**
Pearson correlation coefficient	57.35	55.88	68.38
Spearman correlation coefficient	62.50	59.56	**73.53**
Relief	59.56	61.03	69.12

Moreover, metaheuristic algorithms are applied to both datasets to compare the results. With the same procedure as that used for the filtering algorithms, the same machine learning algorithms are used for the subset of features. As shown in Table 6 and Table 7, the wrapper method based on metaheuristic algorithms is successful in dealing with high-dimensional datasets. The computational cost is high but some of the results are better than those of the filtering algorithms. The results showed that the proposed GA significantly outperformed not only the other metaheuristic algorithms but also the filter-based algorithms. The best accuracy for the Parkinson dataset is 92.07% and great success was achieved at 80.15% for the Cardiac Arrhythmia dataset with RF.

Finally, a hybrid model using filtering and metaheuristic algorithms based on the proposed GA with grid search is implemented. Half of the features are selected with filtering methods and the proposed GA with grid search is applied to a subset of features. The accuracies of all algorithms without feature selection and the proposed GA with grid search are shown in Table 8 and Table 9. According to the results, the best accuracies of the algorithms are achieved with our proposed GA with grid search. Enireddy et al. [11] stated that the best

Table 6. Results of metaheuristic algorithms on Parkinson dataset.

	SVM	**kNN**	**RF**
Proposed Genetic Algorithm	87.67	90.75	**92.07**
Binary Bat Algorithm	85.46	87.22	87.67
Cuckoo Search Algorithm	74.89	80.62	87.22
Equilibrium Optimizer	85.46	87.22	90.31
Genetic Algorithm	84.21	87.50	88.16
Gravitational Search Algorithm	86.34	84.14	88.11
Grey Wolf Optimizer	86.34	83.70	90.75
Harmony Search	85.46	89.87	89.87
Mayfly Algorithm	86.78	89.87	90.31
Particle Swarm Optimizer	85.02	87.67	88.55
Red Deer Algorithm	80.62	85.46	88.11
Sine Cosine Algorithm	85.02	88.55	91.19
Whale Optimization Algorithm	86.78	85.02	90.31

Table 7. Results of metaheuristic algorithms on Cardiac Arrhythmia dataset.

	SVM	**kNN**	**RF**
Proposed Genetic Algorithm	72.79	72.79	**80.15**
Binary Bat Algorithm	59.56	57.35	71.32
Cuckoo Search Algorithm	60.29	59.56	73.53
Equilibrium Optimizer	61.76	60.29	69.12
Genetic Algorithm	58.08	58.09	69.85
Gravitational Search Algorithm	58.82	56.62	68.38
Grey Wolf Optimizer	58.82	57.35	69.12
Harmony Search	60.29	60.29	73.53
Mayfly Algorithm	58.08	58.82	72.79
Particle Swarm Optimizer	61.03	59.56	73.53
Red Deer Algorithm	62.50	58.09	65.44
Sine Cosine Algorithm	59.56	58.82	61.76
Whale Optimization Algorithm	58.82	59.56	65.44

accuracy was 97.96% using GWO with the XGB classifier on the Parkinson dataset. Our remarkable results included 98.24% accuracy with the proposed GA with grid search with the kNN classifier for the same dataset. Contrary to the previous study, 10-fold cross-validation was also used and the results showed that the best accuracies were achieved by the kNN classifier at 80.55%, SVM classifier at 85.85%, and RF classifier at 83.60% based on the proposed GA with grid search.

In addition, Guvenir et al. [3] obtained 62% accuracy using the VFI5 algorithm on the Cardiac Arrhythmia dataset and increased the accuracy to 68% using weights that were learned by a GA. Our best accuracy for cardiac arrhythmia disease classification was 81.62% using the SCC and the proposed GA with grid search with the RF classifier. The previous authors [3] mentioned that they used 10-fold cross-validation to measure the accuracy. To follow the same set-up, we repeated our experiment only for the proposed model and obtained 72.59% accuracy on the same dataset. The promising results showed that the proposed solution has better performance than previous studies. We preferred to use accuracy as an evaluation metric and applied a similar set-up as used in comparable studies. Moreover, we measured the F-score of only the proposed GA with grid search model on both datasets. The results showed that kNN has significant performance of 96.19% on the Parkinson dataset and RF has the best performance among others of 75% on the Cardiac Arrhythmia dataset.

Furthermore, all experiments were done using a 2.4 GHz 4 Core Intel Core i5 processor. Table 10 shows the computational cost in seconds for only the proposed GA and indicates that processing of the Parkinson dataset takes longer than the Cardiac Arrhythmia dataset as expected due to the size of the dataset. Another result demonstrates that RF takes longer to run among other classification algorithms because of the operational volume, such as the maximum depth of the trees.

Table 8. Results of all algorithms on Parkinson dataset.

	SVM	**kNN**	**RF**
Without Feature Selection	85.90	88.00	86.78
Proposed Genetic Algorithm w/Grid Search	94.71	**98.24**	90.75
Binary Bat Algorithm	85.46	87.22	87.67
Cuckoo Search Algorithm	74.89	80.62	87.22
Equilibrium Optimizer	85.46	87.22	90.31
Genetic Algorithm	84.21	87.50	88.16
Gravitational Search Algorithm	86.34	84.14	88.11
Grey Wolf Optimizer	86.34	83.70	90.75
Harmony Search	85.46	89.87	89.87
Mayfly Algorithm	86.78	89.87	90.31
Particle Swarm Optimizer	85.02	87.67	88.55
Red Deer Algorithm	80.62	85.46	88.11
Sine Cosine Algorithm	85.02	88.55	91.19
Whale Optimization Algorithm	86.78	85.02	90.31
Pearson Correlation Coefficient	88.55	84.58	91.19
Spearman Correlation Coefficient	84.14	87.67	86.34
Relief	85.90	86.78	90.75
PCC + Proposed Genetic Algorithm w/Grid Search	91.19	93.83	93.39
SCC + Proposed Genetic Algorithm w/Grid Search	86.78	92.07	91.63
Relief + Proposed Genetic Algorithm w/Grid Search	88.99	89.43	93.39

5. Conclusion and future work

In this study, we used different feature selection strategies to eliminate noise or redundant features. For this purpose, we used filter-based, wrapper, and hybrid methods on the Parkinson and Cardiac Arrhythmia disease datasets and compared the performances. For the wrapper method, we proposed a GA with some specifications. We used three classification algorithms, namely kNN, SVM, and RF, to address the problems of disease classification. The results showed that some metaheuristic algorithms outperformed the filter-based methods in terms of classification accuracy. Another outcome was that the algorithms based on our proposed GA gave better results among other metaheuristic algorithms as well as filter-based algorithms. The proposed GA with grid search with kNN on the Parkinson dataset had 98.24% accuracy and the SCC + proposed GA with grid search as a hybrid method had 81.62% with RF on the Cardiac Arrhythmia dataset. For further analysis, other classification algorithms and their hyperparameter tuning can be used. The models can also be applied to different disease datasets, such as breast cancer or lung cancer.

Table 9. Results of all algorithms on Cardiac Arrhythmia dataset.

	SVM	**kNN**	**RF**
Without Feature Selection	61.03	58.82	72.06
Proposed Genetic Algorithm w/Grid Search	78.68	65.44	77.21
Binary Bat Algorithm	59.56	57.35	71.32
Cuckoo Search Algorithm	60.29	59.56	73.53
Equilibrium Optimizer	61.76	60.29	69.12
Genetic Algorithm	58.08	58.09	69.85
Gravitational Search Algorithm	58.82	56.62	68.38
Grey Wolf Optimizer	58.82	57.35	69.12
Harmony Search	60.29	60.29	73.53
Mayfly Algorithm	58.08	58.82	72.79
Particle Swarm Optimizer	61.03	59.56	73.53
Red Deer Algorithm	62.50	58.09	65.44
Sine Cosine Algorithm	59.56	58.82	61.76
Whale Optimization Algorithm	58.82	59.56	65.44
Pearson Correlation Coefficient	57.35	55.88	68.38
Spearman Correlation Coefficient	62.50	59.56	73.53
Relief	59.56	61.03	69.12
PCC + Proposed Genetic Algorithm w/Grid Search	63.24	67.65	77.94
SCC + Proposed Genetic Algorithm w/Grid Search	72.06	63.97	**81.62**
Relief + Proposed Genetic Algorithm w/Grid Search	66.91	63.97	77.21

Table 10. Collapsed time for proposed GA in seconds.

	Parkinson	Cardiac Arrhythmia
kNN	43.26	12.61
SVM	45.37	13.81
RF	320.57	104.83

References

[1] Balestrino R, Schapira A. Parkinson disease. European Journal of Neurology 2020; 27 (1): 27-42. https://doi.org/10.1111/ene.14108

[2] Sakar CO, Serbes G, Gunduz A, Tunc HC, Nizam H et al. A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable Q-factor wavelet transform. Applied Soft Computing 2019; 74: 255-263. https://doi.org/10.1016/j.asoc.2018.10.022

[3] Guvenir HA, Acar B, Demiroz G, Cekin A. A supervised machine learning algorithm for arrhythmia analysis. In: Computers in Cardiology 1997; Lund, Sweden; 1997. pp. 433-436.

[4] Nemati S, Basiri ME, Ghasem-Aghaee N, Aghdam MH. A novel ACO-GA hybrid algorithm for feature selection in protein function prediction. Expert Systems with Applications 2009; 36 (10): 12086-12094. https://doi.org/10.1016/j.eswa.2009.04.023

[5] Guyon I, Gunn S, Nikravesh M, Zadeh, LA. Feature Extraction: Foundations and Applications. Berlin, Germany: Springer, 2008.

[6] Abualigah L, Mohammad Q. Feature Selection and Enhanced Krill Herd Algorithm for Text Document Clustering. Berlin, Germany: Springer, 2019.

[7] Abualigah LM, Khader AT, Al-Betar MA, Alyasseri ZAA, Essam S. Feature selection with β-hill climbing search for text clustering application. In: 2017 Palestinian International Conference on Information and Communication Technology; Gaza, Palestine; 2017. pp. 22-27.

[8] Saeys Y, Abeel T, Van de Peer Y. Robust feature selection using ensemble feature selection techniques. Lecture Notes in Computer Science 2008; 5212: 313-325. https://doi.org/10.1007/978-3-540-87481-2_21

[9] Hashemi A, Dowlatshahi MB, Nezamabadi-Pour H. A Pareto-based ensemble of feature selection algorithms. Expert Systems with Applications 2021; 180: 115130. https://doi.org/10.1016/j.eswa.2021.115130

[10] Hashemi A, Dowlatshahi MB, Nezamabadi-Pour H. Ensemble of feature selection algorithms: a multi-criteria decision-making approach. International Journal of Machine Learning and Cybernetics 2022; 13: 49-69. https://doi.org/10.1007/s13042-021-01347-z

[11] Enireddy V, Gunda K, Kalyani NL, Prakash KB. Nature inspired binary grey wolf optimizer for feature selection in the detection of neurodegenerative (Parkinson) disease. International Journal of Advanced Trends in Computer Science and Engineering 2020; 9 (3): 3977-3987. https://doi.org/10.30534/ijatcse/2020/222932020

[12] Agrawal P, Abutarboush H, Ganesh T, Mohamed AW. Metaheuristic algorithms on feature selection: a survey of one decade of research. IEEE Access 2021; 9: 26766-26791. https://doi.org/10.1109/ACCESS.2021.3056407

[13] Dökeroğlu T, Deniz A, Kızılöz H. A comprehensive survey on recent metaheuristics for feature selection. Neurocomputing 2022; 494: 269-296. https://doi.org/10.1016/j.neucom.2022.04.083

[14] Diao R, Shen Q. Nature inspired feature selection meta-heuristics. Artificial Intelligence Review 2015; 44: 311-340. https://doi.org/10.1007/s10462-015-9428-8

[15] Oreski S, Oreski G. Genetic algorithm-based heuristic for feature selection in credit risk assessment. Expert Systems with Applications 2014; 41 (4): 2052-2064. https://doi.org/10.1016/j.eswa.2013.09.004

[16] Alyasiri OM, Cheah Y, Abasi, AK, Al-Janabi OM. Wrapper and hybrid feature selection methods using metaheuristic algorithms for English text classification: a systematic review. IEEE Access 2022; 10: 39833-39852. https://doi.org/10.1109/ACCESS.2022.3165814

[17] Abdollahi J, Nouri-Moghaddam B. A hybrid method for heart disease diagnosis utilizing feature selection based ensemble classifier model generation. Iran Journal of Computer Science 2022; 5: 229-246. https://doi.org/10.1007/s42044-022-00104-x

[18] Abd El-Fattah Sayed S, Nabil E, Badr A. A binary clonal flower pollination algorithm for feature selection. Pattern Recognition Letters 2016; 77: 21-27. https://doi.org/10.1016/j.patrec.2016.03.014

[19] Arora S, Singh H, Sharma M, Sharma S, Anand P. A new hybrid algorithm based on grey wolf optimization and crow search algorithm for unconstrained function optimization and feature selection. IEEE Access 2019; 7: 26343-26361. https://doi.org/10.1109/ACCESS.2019.2897325

[20] Karimi F, Dowlatshahi MB, Hashemi A. SemiACO: A semi-supervised feature selection based on ant colony optimization. Expert Systems with Applications 2022; 214: 119130. https://doi.org/10.1016/j.eswa.2022.119130

[21] Hancer E, Xue B, Zhang M, Karaboga D, Akay B. Pareto front feature selection based on artificial bee colony optimization. Information Sciences 2018; 422: 462-479. https://doi.org/10.1016/j.ins.2017.09.028

[22] Al-Tashi Q, Abdulkadir SJ, Rais HM, Mirjalili S, Alhussian H et al. Binary multi-objective grey wolf optimizer for feature selection in classification. IEEE Access 2020; 8: 106247-106263. https://doi.org/10.1109/ACCESS.2020.3000040

[23] Nouri-Moghaddam B, Ghazanfari M, Fathian M. A novel multi-objective forest optimization algorithm for wrapper feature selection. Expert Systems with Applications 2021; 175: 114737. https://doi.org/10.1016/j.eswa.2021.114737

[24] Zhang Y, Zhang Y, Zhang C, Zhou C. Multiobjective Harris hawks optimization with associative learning and chaotic local search for feature selection. IEEE Access 2022; 10: 72973-72987. https://doi.org/10.1109/ACCESS.2022.3189476

[25] Nouri-Moghaddam B, Ghazanfari M, Fathian M. A novel filter-wrapper hybrid gene selection approach for microarray data based on multi-objective forest optimization algorithm. Decision Science Letters 2020; 9: 271-290. https://doi.org/10.5267/j.dsl.2020.5.006

[26] Qin X, Zhang S, Yin D, Chen D, Dong X. Two-stage feature selection for classification of gene expression data based on an improved salp swarm algorithm. Mathematical Biosciences and Engineering 2022; 19 (12): 13747-13781. https://doi.org/10.3934/mbe.2022641

[27] Yang XS. A new metaheuristic bat-inspired algorithm. In: Gonzalez JR, Pelta DA, Cruz C, Terrazas G, Krasnogor N (editors). Nature Inspired Cooperative Strategies for Optimization. Berlin, Germany: Springer, 2010, pp. 65-74. https://doi.org/10.1007/978-3-642-12538-6_6

[28] Yang XS, Deb S. Cuckoo search via Lévy flights. In: World Congress on Nature & Biologically Inspired Computing; Coimbatore, India; 2009. pp. 210-214. https://doi.org/10.1109/NABIC.2009.5393690

[29] Gao Y, Zhou Y, Luo Q. An efficient binary equilibrium optimizer algorithm for feature selection. IEEE Access 2020; 8: 140936-140963. https://doi.org/10.1109/ACCESS.2020.3013617

[30] Holland JH. Genetic algorithms. Scientific American 1992; 267 (1): 66-73.

[31] Rashedi E, Nezamabadi-Pour H, Saryazdi S. GSA: A gravitational search algorithm. Information Science 2009; 179 (13): 2232-2248. https://doi.org/10.1016/j.ins.2009.03.004

[32] Mirjalili S, Mirjalili SM, Lewis A. Grey wolf optimizer. Advances in Engineering Software 2014; 69: 46-61. https://doi.org/10.1016/j.advengsoft.2013.12.007

[33] Geem ZW, Kim JH, Loganathan GV. A new heuristic optimization algorithm: harmony search. Simulation 2001; 76 (2): 60-68. https://doi.org/10.1177/003754970107600201

[34] Bhattacharyya T, Chatterjee B, Singh PK, Yoon JH, Geem ZW et al. Mayfly in harmony: a new hybrid meta-heuristic feature selection algorithm. IEEE Access 2020; 8: 195929-195945. https://doi.org/10.1109/ACCESS.2020.3031718

[35] Han F, Chen WT, Ling QH, Han H. Multi-objective particle swarm optimization with adaptive strategies for feature selection. Swarm and Evolutionary Computation 2021; 62: 100847. https://doi.org/10.1016/j.swevo.2021.100847

[36] Fathollahi-Fard AM, Hajiaghaei-Keshteli M, Tavakkoli-Moghaddam R. Red deer algorithm (RDA): a new nature-inspired meta-heuristic. Soft Computing 2020; 24 (19): 14637-14665. https://doi.org/10.1007/s00500-020-04812-z

[37] Mirjalili S. SCA: A sine cosine algorithm for solving optimization problems. Knowledge-Based Systems 2016; 96: 120-133. https://doi.org/10.1016/j.knosys.2015.12.022

[38] Mirjalili S, Lewis A. The whale optimization algorithm. Advances in Engineering Software 2016; 95: 51-67. https://doi.org/10.1016/j.advengsoft.2016.01.008

[39] Guha R, Chatterjee B, Khalid Hassan SK, Ahmed S, Bhattacharyya T et al. Py_FS: A Python package for feature selection using meta-heuristic optimization algorithms. Computational Intelligence in Pattern Recognition 2022; 1349: 495-504. https://doi.org/10.1007/978-981-16-2543-5_42